

# Understanding the advisor–advisee relationship via scholarly data analysis

Jiaying Liu<sup>1</sup> · Tao Tang<sup>2</sup> · Xiangjie Kong<sup>1</sup>  · Amr Tolba<sup>3,4</sup> ·  
Zafer AL-Makhadmeh<sup>3</sup> · Feng Xia<sup>1</sup>

Received: 8 October 2017 / Published online: 4 May 2018  
© Akadémiai Kiadó, Budapest, Hungary 2018

**Abstract** Advisor–advisee relationship is important in academic networks due to its universality and necessity. Despite the increasing desire to analyze the career of newcomers, however, the outcomes of different collaboration patterns between advisors and advisees remain unknown. The purpose of this paper is to find out the correlation between advisors’ academic characteristics and advisees’ academic performance in Computer Science. Employing both quantitative and qualitative analysis, we find that with the increase of advisors’ academic age, advisees’ performance experiences an initial growth, follows a sustaining stage, and finally ends up with a declining trend. We also discover the phenomenon that accomplished advisors can bring up skilled advisees. We explore the conclusion from two aspects: (1) Advisees mentored by advisors with high academic level have better academic performance than the rest; (2) Advisors with high academic level can raise their advisees’ h-index ranking. This work provides new insights on promoting our understanding of the relationship between advisors’ academic characteristics and advisees’ performance, as well as on advisor choosing.

**Keywords** Academic networks · Scholarly data · Social network analysis · Advisor–advisee relationship · Collaboration patterns

---

✉ Xiangjie Kong  
xjkong@ieee.org

<sup>1</sup> Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, School of Software, Dalian University of Technology, Dalian, China

<sup>2</sup> Chengdu College, University of Electronic Science and Technology of China, Chengdu, China

<sup>3</sup> Computer Science Department, Community College, King Saud University, Riyadh, Saudi Arabia

<sup>4</sup> Mathematics and Computer Science Department, Faculty of Science, Menoufia University, Shibin El Kom, Egypt

## Introduction

Scholarly data analytics has become a vital part of the scientific research (Letchford et al. 2015; Lehmann et al. 2006; Glänzel et al. 2017). Relationship extraction is a major area of interest in the scholarly social network analysis. The structure of the network usually contains a large number of nodes, with different relationships such as friendship, kinship, and hostility. Analyzing such relations can strengthen our understanding of the evolution and development of the scholarly society (Xia et al. 2017). One of the main relations is the mentorship because mentors are important for their protégés. Researchers have studied various aspects of mentorship including types, phases, outcomes and mobilities as early as 30 years ago (Chao et al. 1992; Kram 1983). Although there is little doubt that the study of academic mentorship is necessary, it has been neglected because there is no complete advisor–advisee dataset. Each scholar needs guidance from the advisor when she/he enters academia as a newcomer. It is widely accepted that different advisors influence advisees differently (Murphy 2015). However, where these differences are originated and how significant they are, both are worth exploring. The lack of quantitative academic mentorship analysis leads to the striking lack of specific guiding significance for advisors choosing.

The mentorship is universal and complex. Traditionally, a mentorship is defined as “A relationship between the inexperienced mentee and an experienced senior member of a particular field” (Dobson 2013). It has different types and forms in different fields. For instance, it can be regarded as the relationship between employees and managers in public utility companies and advisor–advisee relationship in academia. There is a large volume of studies describing mentorship in various aspects. According to the formation of the relationship, Chao et al. (1992) first categorized mentorship into two types: formal mentorship and informal mentorship. Informal mentorship usually has neither compact structure with reasonable management, nor been recognized by the organization formally. In contrast to informal mentorship, formal mentorship occurs with external intervention from the organization. Usually it is led and managed by the organization. Mentorship is a temporary relationship with almost fixed duration, for example, the average length of advisor–advisee relationship in academic networks is five years (Kram 1983). The phases of the mentorship can be defined as initiation, cultivation, separation, and redefinition on the basis of the particularly effective experience (Kram 1983). In recent years, there has been an increasing amount of literature describing specific elements of mentorship, including behaviors of mentors and protégés, mentoring functions, structure of cooperation in the mentorship (Johnson and Ridley 2015; Bozionelos et al. 2014; Borders et al. 2012), and so on. These studies provide complete insights into the development and implementation of the mentorship.

No matter what the form is, the ultimate goal of mentorship is to enhance the personal and professional development of protégés (Chao 1997). Therefore, mentors have been taken as beacon lights in the development of protégés. Based on inherently dyadic attributes of the mentorship, the existing literature demonstrates that mentorship is positive and beneficial for both mentors and protégés. In a positive mentorship, the protégés will have better performance and more opportunities for promotion because their mentors provide them with personal and career assistance. Meanwhile, the mentors can receive fulfillment and gain satisfaction by improving protégés’ welfare (Ghosh and Reio 2013; Hu et al. 2014). Furthermore, organizations benefit as well because there is a real possibility that protégés would like to be devoted to their organizations after graduation (Malmgren et al. 2010; Florea et al. 2013). Moreover, as early as 1989, Fagenson (1989) suggested that

mentoring can enhance work validity and the career success. Singh et al. (2009) perceived to the point that compared with those who do not have a mentor, individuals who have obtain mentors embrace more chances to promote in their careers, and more likely to consider their future prospects ahead of time in more positive ways. Wanberg et al. (2006) proposed sometimes a mentor may change a protégé's vocational tendency which can lead the protégé to the different lifeline.

The recent trends in mentorship analysis have led to a proliferation of studies focusing on the association between mentorship and the career development of protégés. Since Kram (1983) presented the psychosocial functions and career development functions to examine the mentors' influence on the protégés, almost every paper focusing on the outcome of mentorship includes contents related to it. The mentoring functions have received considerable critical attention. On this basis, Chao et al. (1992) compared the mentoring functions among formal mentorships, informal mentorships, and non-mentored counterparts. Scandura and Ragins (1993) investigated the link between mentoring functions and the career mobility outcomes in terms of promotions and salaries. In a study conducted by Singh et al. (2009), it showed the correlation of rising star attributes measured at different periods, i.e., no-mentored period and a year after mentored. In the same vein, Chao (1997) examined linkages among different mentorship phases, functions and outcomes of protégés in order to integrate different aspects of mentoring into a more comprehensive theory. According to Young and Perrewé (2000), there were specific behaviors such as trust, related to career and social support the exhibited throughout the mentoring process. Furthermore, Tuesta et al. (2015) found the evidence that there exists a positive correlation between time of advisor–advisee relationship and advisee's productivity in the area of Exact and Earth Sciences.

Although research has been carried out on the mentors' influence, with the exception of Tuesta et al. (2015) who measured scientific productivity by the number of publications in journals, all others (Chao et al. 1992; Scandura and Ragins 1993; Singh et al. 2009; Young and Perrewé 2000; Chao 1997) remain narrowly focus on dealing only with analytic study based on the feedback from the questionnaires. There are few data-based investigations studying the outcomes of choosing different advisors in academia. This gives rise to the question of how to choose an advisor when newcomers enter academia. In response to the question “What kind of advisors you will choose when you decide to pursue the PhD degree?”, most of the answers are non-specific and all-embracing. Almost all advisees want to choose advisors who make a great and positive influence on themselves. Young advisees in academia often seek to work in collaboration with top advisors in their field in pursuit of a successful career. However, “top advisors” in academia can be defined differently.

In this work, we analyze the relationship between advisees' performance (i.e., productivity and impact) and advisors' academic characteristics from the perspective of Computer Science. It is different from the existing research which only employs questionnaires. The dataset of advisor–advisee relationships is extracted from Digital Bibliography & Library Project (DBLP) dataset (Ley 2009). We use the improved stacked autoencoder method based on the deep learning with the highest accuracy reaching up to 94% to get the complete dataset (Wang et al. 2016, 2017). Then we make a quantitative analysis of the dataset. We calculate the productivity and impact of advisees including the number of publications, citations, and h-indices in 1–12 years after first collaborating with their advisors. We separate the advisors into groups according to their academic age to observe the relation between advisees' outcomes and advisors' academic ages. We find that advisees' number of publications, citations, and h-indices follow the same trend with the

increasing of advisors' academic ages, exhibiting an initial growth, remaining stationary for a duration, finally with a declining trend. Based on the observation, we examine the correlation between the ranking of advisors and their advisees' h-indices.

Furthermore, by combining quantitative analysis with qualitative analysis, we find that advisors with high academic levels will bring up advisees with high academic performance. We conclude the findings from two main aspects: (1) Advisees mentored by the advisors with high academic level generally have better academic performance than the rest. We divide the advisors into two groups: one with a high ranking in terms of publications/citations/h-indices and the other without. By comparing the academic performance of all advisees mentored by different advisors, we observe that the phenomenon also holds for advisees with high academic level (if a scholar ranks the top 10% in the number of publications/citations/h-indices, then she/he has the higher academic achievement). (2) Advisors with high academic level can increase the probability of their advisees ranking the top 10% in h-index. We calculate the probability of an advisee ranking the top 10% in different cases. These cases are based on advisors' different h-indices and academic ages. The outcomes of this quantitative research, which compare advisees' achievements coached by advisors with the different academic performance we present here, can be utilized for advisor recommendation.

## Methods

### Dataset

Our large-scale dataset of advisor–advisee relationships is generated by applying a deep learning technique, stacked autoencoder, on the DBLP dataset (Wang et al. 2016, 2017). The specific steps for the advisor–advisee dataset acquisition are as follows:

1. Construct the collaboration ego-network for all scientists in the DBLP dataset.
2. Clean the network: Following Sinatra's (Sinatra et al. 2016) steps, we only choose the scholars who have published at least 5 papers and who have a publication career span of at least 10 years in the DBLP dataset.
3. Extract the required features for training as unlabeled input of the model: Compute the personal properties (i.e., academic age, number of publications) and collaboration properties (i.e., collaboration times, collaboration duration, times of first-two authors, and cohesion of collaboration) for each scholar in the first 8 years of collaboration, then use these normalized features as the input of the model. The Back Propagation (BP) method is used to train the model and optimize the model. The result of identifying advisors is obtained through classifier after training.
4. Select the advisor–advisee pair whose identification accuracy exceeds 90% as the advisor–advisee dataset.

The output dataset consists of 810,469 advisor–advisee pairs and more than 1,475,000 publications from 1968 to 2016, with the highest computational precision reaching 94%. While DBLP only provides publication details for each scholar, the accomplishment in terms of citations cannot be obtained from DBLP. In order to obtain scholars' citations and h-indices, we match our dataset with AMiner-DBLP (Tang et al. 2008b). AMiner<sup>1</sup> is a website which aims to provide comprehensive search and mining services for researcher

---

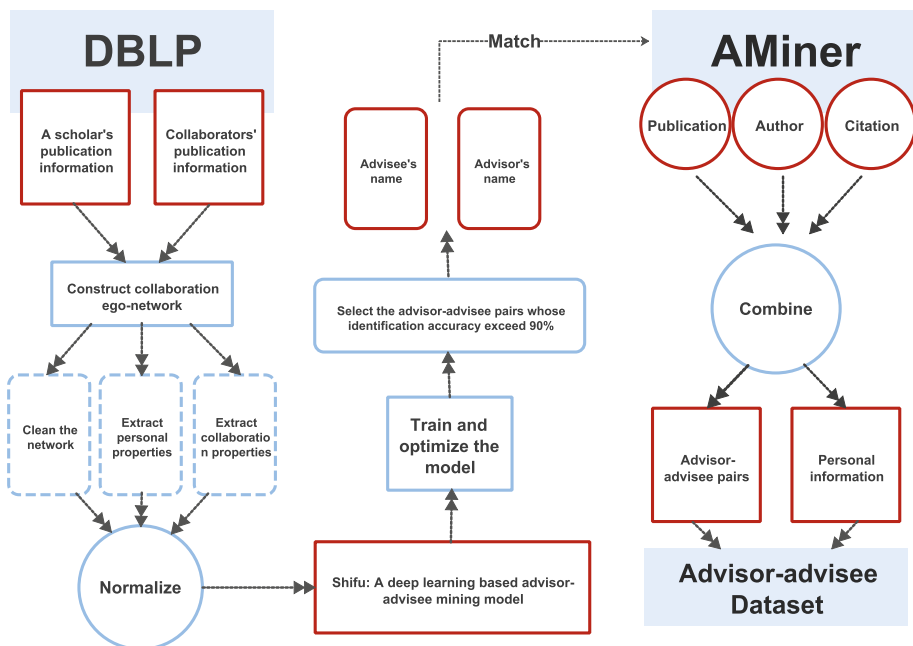
<sup>1</sup> <https://www.aminer.cn>.

social networks. It is considered as a widely used and one of the best-curated databases for Computer Science articles (Gollapalli et al. 2011; Moreira et al. 2011; Tang et al. 2008a; Amjad et al. 2017). Currently, the system consists of more than 6000 conferences, 3,200,000 publications, and 700,000 researcher profiles before 2016. In this website, developers use Microsoft Graph Search API to query each AMiner paper's title and obtain candidate matching papers for each AMiner paper. For the computational accuracy, they randomly sampled 100,000 linking pairs and evaluated the matching accuracy. The number of truly matching pairs is 99,699 and the matching accuracy can achieve 99.70%.

Finally, the analysis is on the grounds of 15,559 advisors whose academic age is 5 (23,473 advisees), 20,859 advisors whose academic age is 10 (36,841 advisees), 17,028 advisors whose academic age is 15 (33,652 advisees), 11,522 advisors whose academic age is 20 (24,883 advisees), and 7352 advisors whose academic age is 25 (11,305 advisees). The whole process of the experiment is illustrated in Fig. 1.

## Ranking the scholars

We take three variables into consideration to rank a scholar: number of publications, citations, and h-indices. H-index of a scholar is calculated on the basis of the year of publications and referenced information each year. A scholar has an h-index  $h$  if at least  $h$  of his/her publications attract  $h$  or more citations. In order to measure the advisors' influence on advisees during different collaboration periods, the indicators for each year are calculated.



**Fig. 1** Dataset acquisition process

## Analysis method

### *Academic age of scholars*

When dividing the type of advisors, we take advisors' academic ages into consideration because the academic characteristics such as citations and the number of publications are accumulated over time. Therefore, advisors with different academic ages should not be put together for comparison.

**Definition 1** AA is defined as the academic age of scholars, i.e.,

$$AA = Y_c - Y_f$$

where  $Y_f$  is the year scholars published the first article and  $Y_c$  is the investigated year.

In this paper, we calculate advisees' and advisors' academic ages of the year in which they began cooperating with each other. In other words,  $Y_c$  is the year when they commenced collaboration.

### *The probability of advisees whose h-indices rank the top 10%*

In order to observe the association between advisees' academic performance and advisors' academic level, we calculate the possibility that advisees mentored by different types of advisors with different h-index rankings and academic ages. The steps are described as follows:

- (1) Divide the advisors according to the academic age when they first collaborated with their advisees. For advisors with the same academic age, divide them into Top<sub>10</sub> group in which advisors' h-index ranking is in the top 10% and *Res* group for the rest.
- (2) Calculate the total number of Top<sub>10</sub> advisors' advisees  $N_{t_i}$  and *Res* advisors' advisees  $N_{r_i}$ .
- (3) Rank all the advisees with the same academic age according to their h-indices.
- (4) Calculate the minimum of the top 10% advisees' h-index  $h_{\min}$ .
- (5) Calculate the number of Top<sub>10</sub> advisors' advisees  $Nh_{t_i}$  whose h-indices are above  $h_{\min}$ . In the same way, calculate  $Nh_{r_i}$  of *Res* advisors.
- (6) The probability of Top<sub>10</sub> advisors' advisees whose h-index rank top 10% can be calculated as  $Nh_{t_i}/N_{t_i}$  and  $Nh_{r_i}/N_{r_i}$  for *Res* advisors.

## Results

In this section, we present the basic analysis of the advisor–advisee dataset in “[Statistical analysis of the advisor–advisee relationships](#)” Section. “[Correlation between advisors' academic characteristics and advisees' academic performance](#)” Section elaborates the correlation between advisors' academic characteristics and advisees' academic performance. “[An accomplished advisor brings up a skilled advisee](#)” Section shows the phenomenon that an accomplished advisor could bring up a skilled advisee. Finally, the analysis of the increasing publication rate is highlighted in “[Publication rates analysis](#)” Section.

## Statistical analysis of the advisor–advisee relationships

Table 1 summarizes the statistics of the advisors at the beginning of collaborating with their advisees. It illustrates the mean of advisors’ academic characteristics including the number of publications (NP), citations (NC), and h-indices (NH). Here we separate advisors according to their exact academic age (AA) when they begin guiding the advisees, to mitigate the bias of accumulated citations for advisors with different AA.

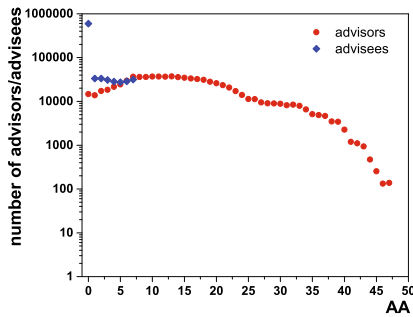
In order to carry out the data-based analysis of the advisor–advisee relationships, we keep a tally of relevant information of advisors and advisees in the dataset. Figure 2 shows the range of each personal characteristic and number of advisors/advisees distribution of these characteristics.

The red points represent advisors and blue ones represent advisees. Figure 2a shows the academic age distribution of advisors and advisees since their first collaboration. As shown in Fig. 2a, most advisors’ AA are 5–20 years. After that, as AA increases, the number of advisors decreases. Figure 2c, e, g, respectively illustrate the distributions of advisees (10 years after first collaborating with their advisors) and advisors (when they began mentoring a certain advisee) in terms of NP, NC, and NH. We only consider the advisees whose career spans at least 10 years since they first collaborated with their advisors. It can be found from Fig. 2c, e, g, that there is a clear decreasing trend in the number of scholars with the increasing of NP, NC, and NH. We also plot the survival function, that is  $P(x > X)$  for these graphs, which will get rid of the noise and observe the differences between various groups. The results are presented in Fig. 2b, d, f, h. In these figures, the horizontal axis represents scholars’ AA, NP, NC, and NH, respectively. The vertical axis is the survival rate. These survival curves have a declining trend. The steeper decline slope is on behalf of the lower survival rate. From these figures, we can see that all curves are concave. It illustrates that most scholars have relatively lower academic performance, few scholars have higher academic achievements, which can be evidenced by “Long Tail Effect” (Anderson 2006).

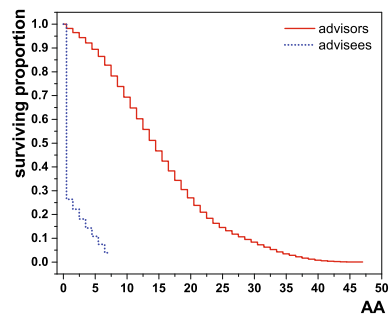
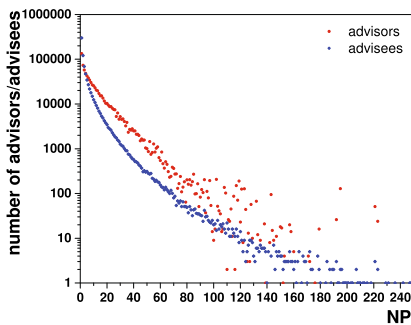
Graphs in Fig. 3 are the results obtained from fitting methods for Fig. 2. Specific fitting functions are summarised in Table 2. In Table 2,  $y$  represents the number of advisees/advisors and  $x$  represents AA/NP/NC/NH. In order to carry out effective fitting, we use different exponential functions to determine the relationship. In fact, the existing literature (Newman 2001) has pointed out that the degree distribution of the scientists’ networks is between the exponential and the power-law. Obviously, our results are in consistent with aforementioned law.

**Table 1** Academic characteristics of advisors

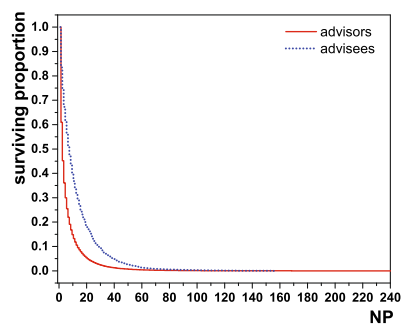
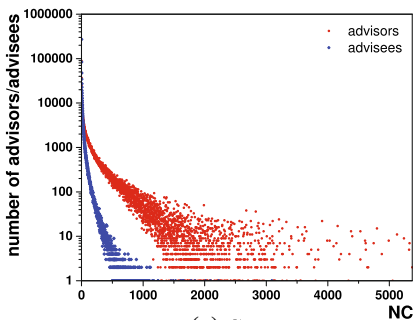
Attributes	AA = 5	AA = 10	AA = 15	AA = 20	AA = 25	Average
NP	12.21	17.65	12.01	10.22	9.91	11.73
NC	57.11	145.2	195.05	272.74	319	160.4
NH	3.13	5.29	5.93	6.84	7.25	4.88



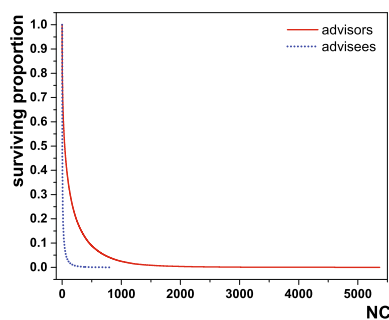
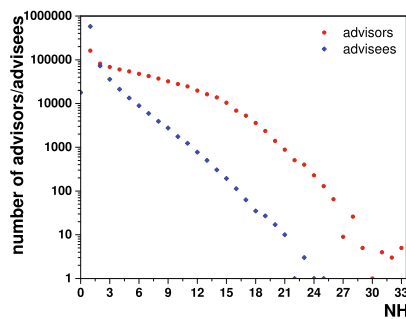
(a) Academic ages

(b) Survival function for *AA*

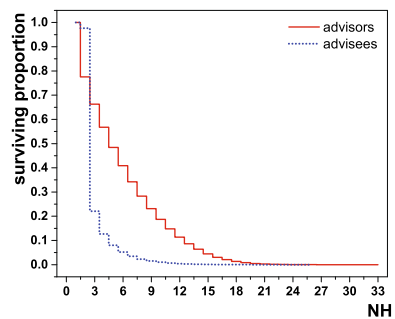
(c) Number of publications

(d) Survival function for *NP*

(e) Citations

(f) Survival function for *NC*

(g) H-indices

(h) Survival function for *NH*



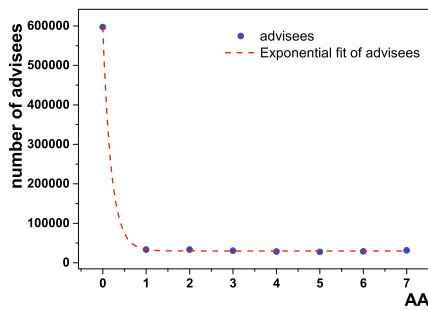
◀ **Fig. 2** Academic performance distribution and survival function for advisors and advisees. The horizontal axis represents **a** academic ages, **c** the number of publications, **e** citations, and **g** h-indices, respectively. The vertical axis represents the number of scholars. In **b**, **d**, **f**, and **h**, the vertical axis represents the surviving proportion

### Correlation between advisors' academic characteristics and advisees' academic performance

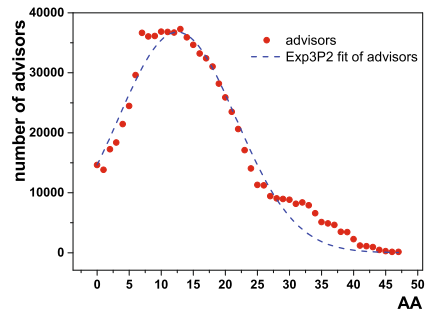
Figure 4 shows the relationship between the advisors' academic characteristics and advisees' academic productivity as well as their impact. Academic productivity can be represented by the number of publications authored by one scholar, and academic impact can be approximated by a scholar's h-index (Azoulay 2012; Owens 2013). It is a comprehensive indicator to evaluate scholars. Figure 4a shows the correlation between advisors' AA and average NP, NC, and NH of their advisees 10 years after first collaborating with their advisors. It can be seen that the indicators have the same trend with the increasing of AA because these indicators have a certain dependence. All of them experience an initial growth, then remain stationary, and finally reach a decline phase. When advisors' academic ages are in the range 1–9, advisees' academic performance shows a linear growth trend. When advisors' academic age reaches 28, advisees' performance shows a downward trend. In the period when advisors' academic age is between 10 and 27, advisees' average achievements reach the highest level and basically remain unchanged.

It is interesting to consider the possible factors underlying this phenomenon. Prior studies have noted the importance of advisors' career coaching and social support in advisees' development. The current study (Wang et al. 2017) finds that the productivity of scholars is dynamic over the course of their own scientific career. For scholars whose  $AA < 12$ , the annual productivity rises up slowly. If their  $AA > 24$ , their productivity will decline slightly with the AA. In the range of  $12 < AA < 24$ , their annual productivity are nearly fixed. Advisees' average productivity/impact shows a linear growth trend because their advisors' productivity in the range  $1 \leq AA \leq 9$  increases linearly. This reflects that the relationship between advisors' career phases and advisees' outcomes in terms of productivity and impact actually exists in the scientific mentorship. Our analyses indicate the relationship between collaborations with different AA advisors and the future career performance of advisees thus need deeper exploration.

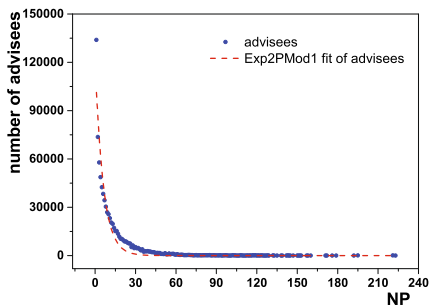
Here are some possible reasons for our findings: (a) Academic research is a long-term accumulation process. Junior advisors need time to accumulate experience in conducting research and mentoring advisees. Compared with senior ones, their capacity and resources are limited. Senior advisors have cumulative advantages, both in academic performance and resources. (b) When the academic age of advisors is at the intermediate level (in the range of 10–27 years), their academic capability may reach a certain level as well. With sufficient resources, they can provide their advisees with the most helpful guidance. In addition to that, they are also willing to strive to instruct their advisees, because their own academic performance can be improved in this way. (c) Few senior advisors have a relatively long academic career (as shown in Fig. 2a), that's why senior advisors of academic age above 35 have poor performing advisees. The random fluctuation may cause this phenomenon, and thus further study is required to untangle the underlying reasons. In general, the correlation between advisees' high performance and advisors' academic ages is certainly reasonable, but it is not the causation of scholars' success. Hard work and passion are still essential to scientists' success in the scientific career.



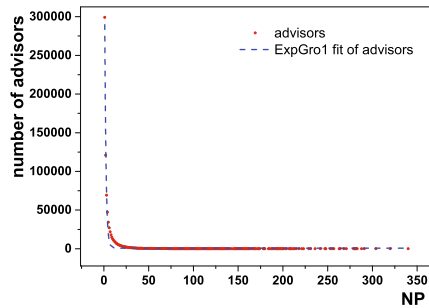
(a) AA of advisees



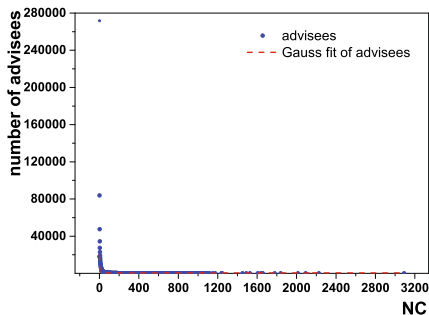
(b) AA of advisors



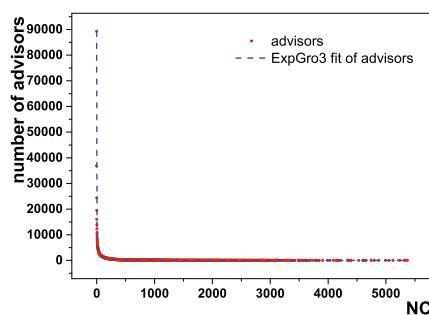
(c) NP of advisees



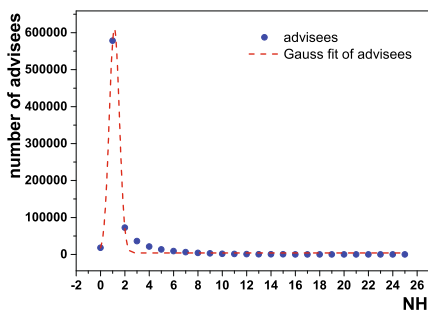
(d) NP of advisors



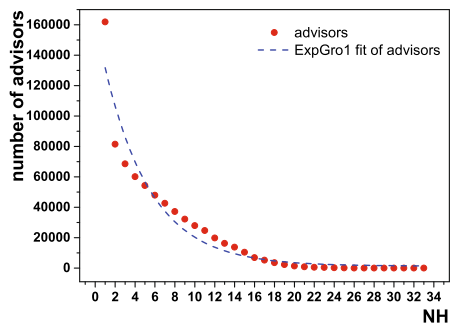
(e) NC of advisees



(f) NC of advisors



(g) NH of advisees

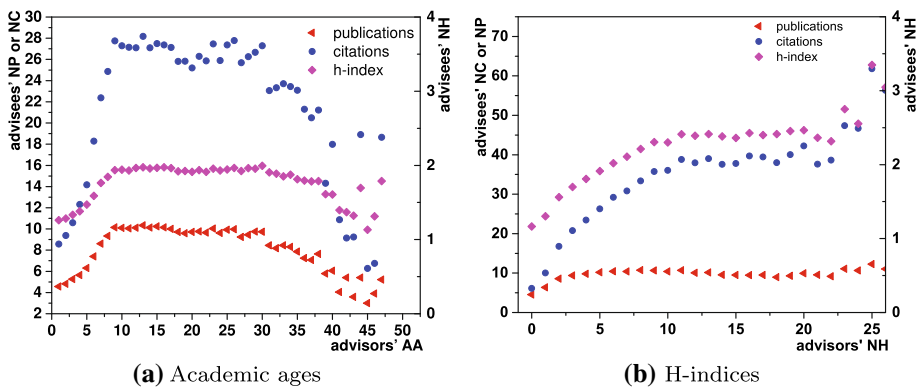


(h) NH of advisors

◀ **Fig. 3** Fitted curves plot for AA, NP, NC, and NH. The horizontal axis represents **a** advisees' AA, **b** advisors' AA, **c** advisees' NP, **d** advisors' NP, **e** advisees' NC, **f** advisors' NC, **g** advisees' NH, and **h** advisors' NH, respectively. The vertical axis represents the number of scholars

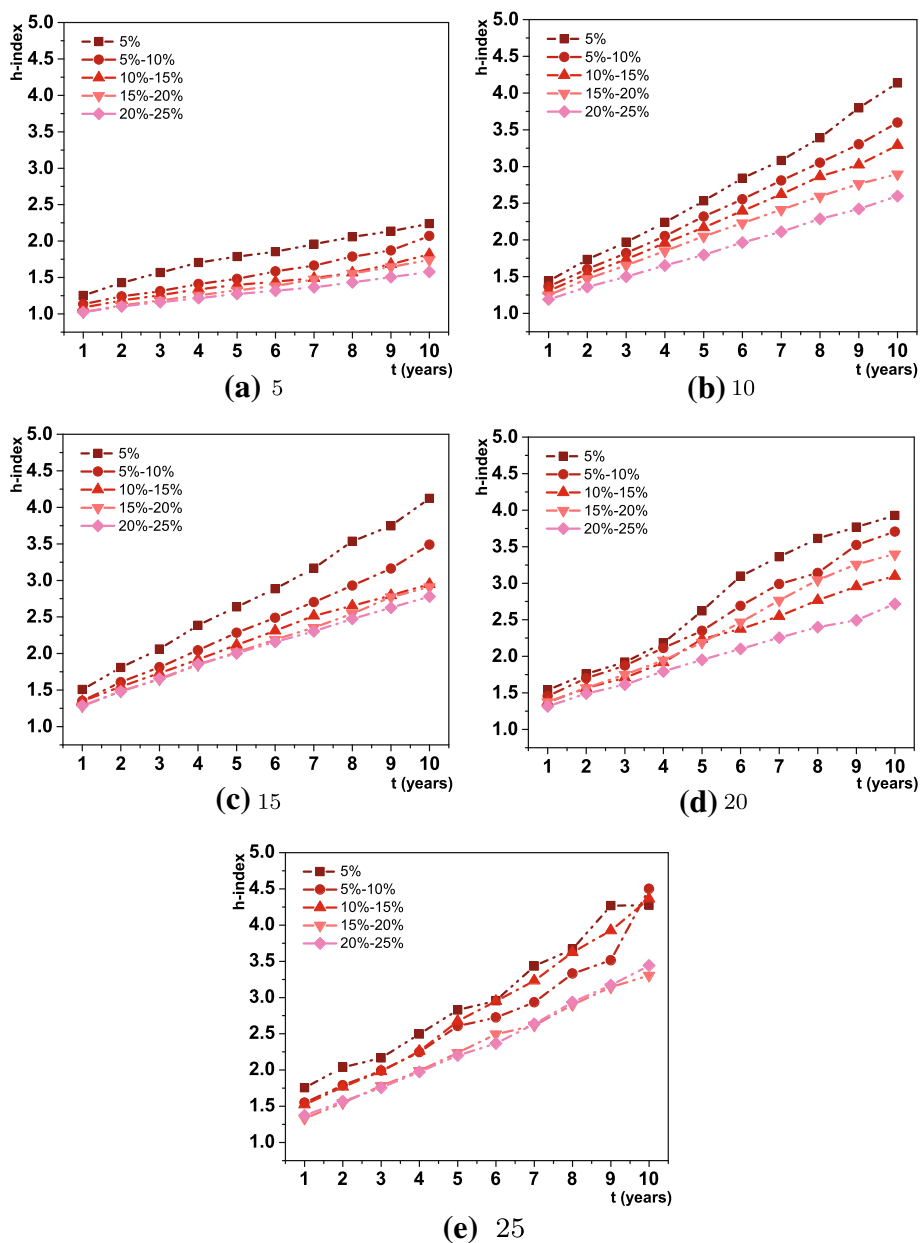
**Table 2** Exponential fitting of AA, NP, NC, and NH distribution for advisors and advisees

Attributes	Fitting function
Advisees' AA	$y = y_0 + Ae^{R_0 x}$
Advisors' AA	$y = Ae^{-AX}$
Advisees' NP	$y = ae^{bx}$
Advisors' NP	$y = y_0 + A_1 e^{\frac{x}{t_1}}$
Advisees' NC	$y = y_0 + \frac{A}{w\sqrt{\frac{\pi}{2}}} e^{-\frac{2(x-x_0)^2}{w^2}}$
Advisors' NC	$y = y_0 + A_1 e^{\frac{x}{t_1}} + A_2 e^{\frac{x}{t_2}} + A_3 e^{\frac{x}{t_3}}$
Advisees' NH	$y = y_0 + \frac{A}{w\sqrt{\frac{\pi}{2}}} e^{-\frac{2(x-x_0)^2}{w^2}}$
Advisors' NH	$y = y_0 + A_1 e^{\frac{x}{t_1}}$



**Fig. 4** Relationship between the advisors' academic characteristics and advisees' academic performance for 10 years since collaborating with their advisors. The horizontal axis represents **a** advisors' academic age, **b** h-indices. In the subgraphs, the left vertical axis represents advisees' average number of publications and citations. The right vertical axis represents their average h-index

Similarly, in Fig. 4b, the abscissa represents the advisors' NH. In this subgraph, the left vertical axis represents advisees' average NP, NC and right axis represents their average NH. Comparing with advisees' NC and NH, advisees' NP has the least obvious trend with the increase of advisors' NH. The phenomenon elicits the truth that quantity is not equal to quality especially in academia. The maximum value of advisors' h-index we have considered is 25 because there are few advisors with high NH values (see Fig. 2g). The value of indicators for advisees' fluctuates widely along with the growth of abscissa. So the data is binned in order to make the trend more clear.



**Fig. 5** Advisees' h-indices whose advisors' academic ages are **a** 5, **b** 10, **c** 15, **d** 20, **e** 25, as well as their h-indices are ranking at the top 5%, 5%–10%, 10%–15%, 15%–20%, 20%–25%. The polylines represent the trends of all advisees' average h-index over time

### An accomplished advisor brings up a skilled advisee

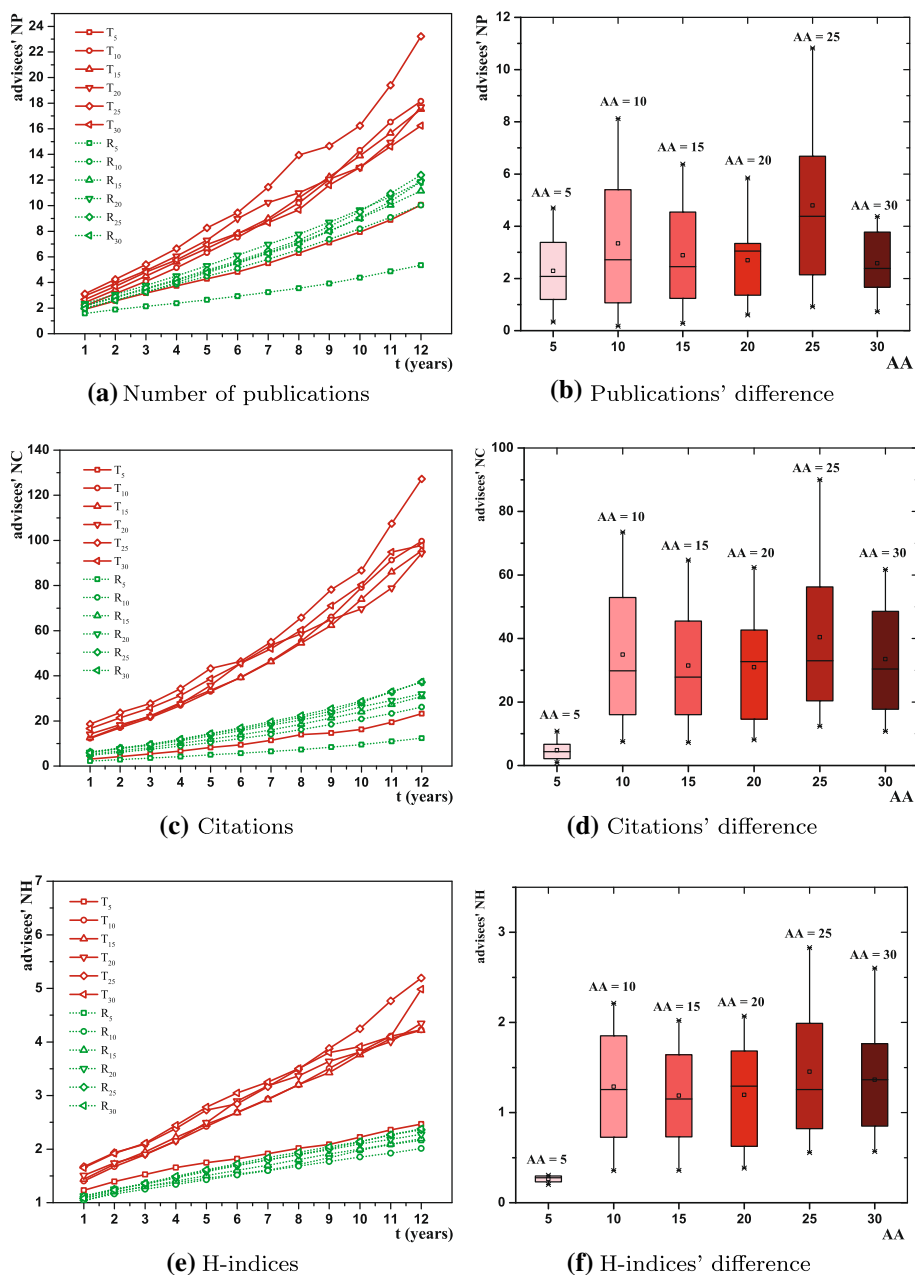
According to the influence of advisors' academic ages on advisees, Fig. 5 indicates advisees' h-indices whose advisors' academic ages are 5, 10, 15, 20, 25, and the ranking of

their h-indices begins at the top 5%, 5%–10%, 10%–15%, 15%–20%, 20–25%, respectively. The polylines represent the trends of all advisees' average h-index over time. It can be seen from the data in Fig. 5, advisors with different rankings in h-index can bring their advisees different h-indices. It is commonly perceived that high-achieving advisors literally cause their advisees higher academic productivity and impact. Selective bias is one possible reason as top advisors tend to select quality students. It is interesting to observe that as the advisors' academic age rises, the gap among advisees narrows.

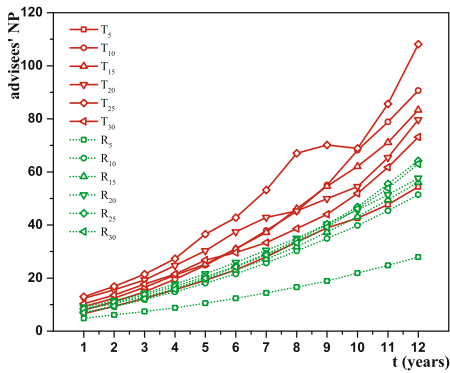
The advisees supervised by advisors whose h-index rank the top 10% always have the highest h-index. But it is not clear how significant the differences among these advisees are. On the basis of the definition of scholars' impact, here we regard the top ten percent of advisors in h-index as  $Top_{10}$  advisors. *Res* advisors represent the rest of advisors. We try to find out the differences between advisees mentored by  $Top_{10}$  advisors or *Res* advisors. We convert the problem from a qualitative analysis to the quantitative analysis. Figure 6 reflects advisees' average NP, NC, and NH over time since advisees first collaborated with their advisors. In Fig. 6, advisors are differentiated with academic ages and h-indices. In the figures,  $T_n$  corresponds to  $Top_{10}$  advisors whose academic ages are  $n$  while  $R_n$  corresponds to *Res* advisors whose academic ages are  $n$ . From the figures, we can see that advisors' NP results in the least obvious discrepancy between advisees supervised by  $Top_{10}$  advisors and *Res* advisors. The most striking observation to emerge from the data comparison is the growth rate of each indicator. Advisees mentored by  $Top_{10}$  advisors have higher NP, NC, and NH than the others. Moreover, their growth rate is higher. This result may be explained by the Matthew Effect (Langfeldt et al. 2015). In academia, while excellent scholars' accomplishments and reputation tend to snowball, those with modest accomplishments have greater difficulty to improve their impact.

In Fig. 7, similarly, which is consistent with assessment standards of advisors, the advisees whose h-index rank the top 10% of each indicator are considered as the  $Top_{10}$  advisees. The differences between  $Top_{10}$  advisees' NP, NC, and NH, who are mentored by  $Top_{10}$  advisors or *Res* advisors are highlighted in Fig. 7d, e, f. To the same extent as the previous results, comparing with NC and NH, the discrepancy between advisees' average NP mentored by  $Top_{10}$  advisors or *Res* advisors is not obvious. However, considering academic ages of the advisors,  $Top_{10}$  advisees mentored by  $Top_{10}$  advisors whose academic age is 30 have the lowest average NH. For *Res* advisors, the same scenario occurs to advisors whose academic ages are 5. It is different from advisors' NC and NH. This phenomenon gives us inspiration that only considering the number of publications to evaluate scholars is insufficient because quantity is not always equal to quality. As a result, a number of comprehensive indices have appeared to evaluate scholars. In Figs. 6 and 7, it can be seen from the subgraphs that comparing with all of the advisees,  $Top_{10}$  advisees mentored by  $Top_{10}$  advisors or *Res* advisors make more distinct values in the average NP, NC, and NH.

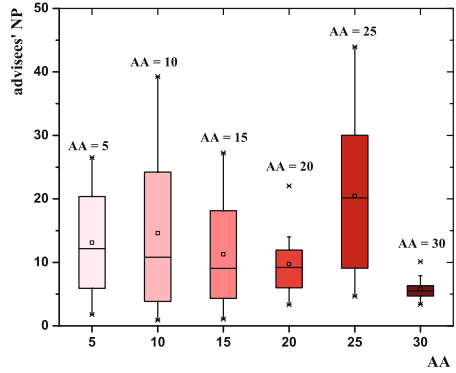
The results of the correlational analysis among advisor–advisee's collaboration duration, advisors' academic ages, proportion of advisees' h-index ranking the top 10% are shown in Fig. 8. The data in first three years of collaboration with advisors whose academic age is 5 are missing because of the slow growth rate in this period. It is difficult to calculate the top 10% advisees' h-indices because the values of their h-indices are relatively low and similar. For *Res* advisors, the proportion of advisees ranking top 10% is shown in the left subgraph in Fig. 8a, and the result of  $Top_{10}$  advisors is summarized in the right subgraph. Figure 8b shows the proportion of differences between advisees mentored by  $Top_{10}$  advisors and *Res* advisors. The values of advisors whose academic ages are 30 in Fig. 8b



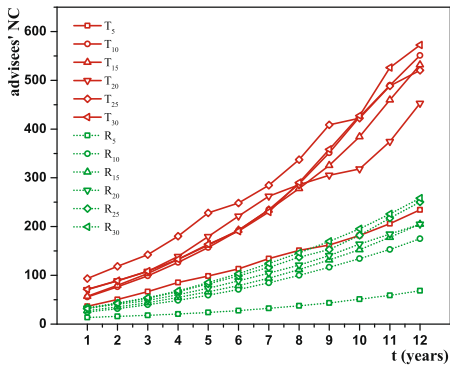
**Fig. 6** Average academic performance of all advisees supervised by Top<sub>10</sub> advisors and Res advisors.  $T_n$  corresponds to Top<sub>10</sub> advisors whose academic ages are  $n$ .  $R_n$  corresponds to the rest whose academic ages are  $n$ . **a**, **c**, and **e** show the advisees' number of publications, citations and h-indices coached by different advisors, respectively. Details of their statistical mean and deviation are shown in **b**, **d**, and **f**, respectively



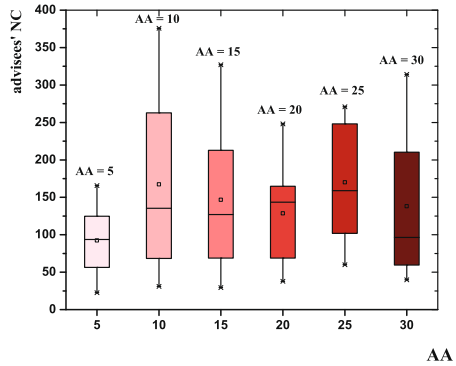
(a) Number of publications



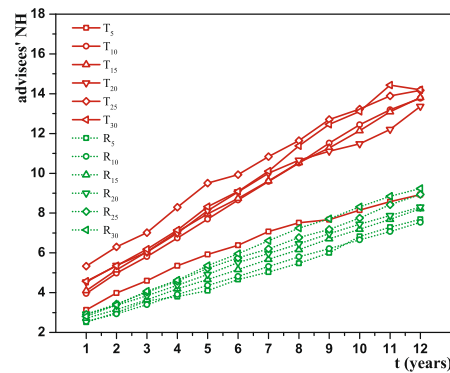
(b) Publications' difference



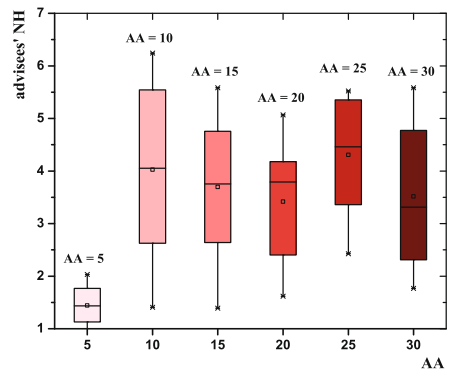
(c) Citations



(d) Citations' difference

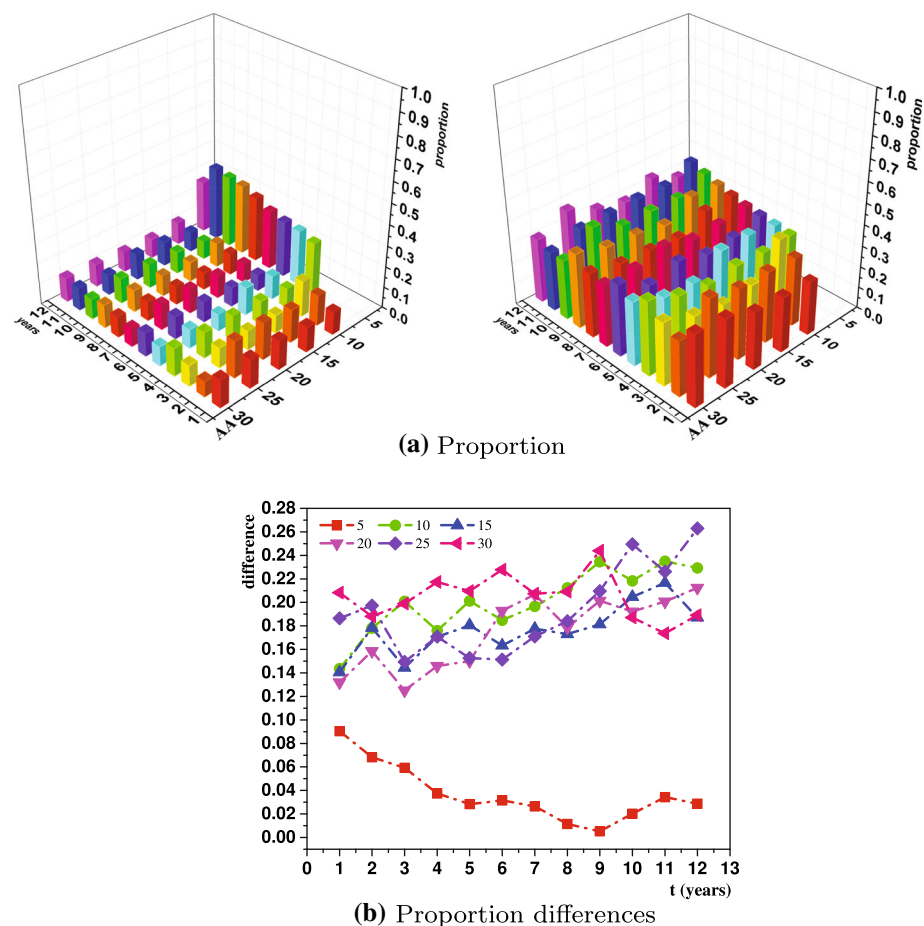


(e) H-indices



(f) H-indices' difference

**Fig. 7** Top<sub>10</sub> advisees' academic performance supervised by Top<sub>10</sub> advisors and *Res* advisors. **a**, **c**, and **e** show the Top<sub>10</sub> advisees' number of publications, citations and h-indices coached by different advisors. Differences between Top<sub>10</sub> advisees' NP, NC, and NH who are mentored by Top<sub>10</sub> advisors or *Res* advisors are highlighted in **b**, **d**, and **f**



**Fig. 8** Correlations among advisor–advisee’s collaboration duration, advisors’ academic age, the proportion of advisees’ h-index ranking the top 10%. **a** Advisees’ probability of h-index ranking the top 10% mentored by *Res* and *Top<sub>10</sub>* advisors. **b** The proportion of differences between advisees mentored by *Top<sub>10</sub>* advisors and *Res* advisors

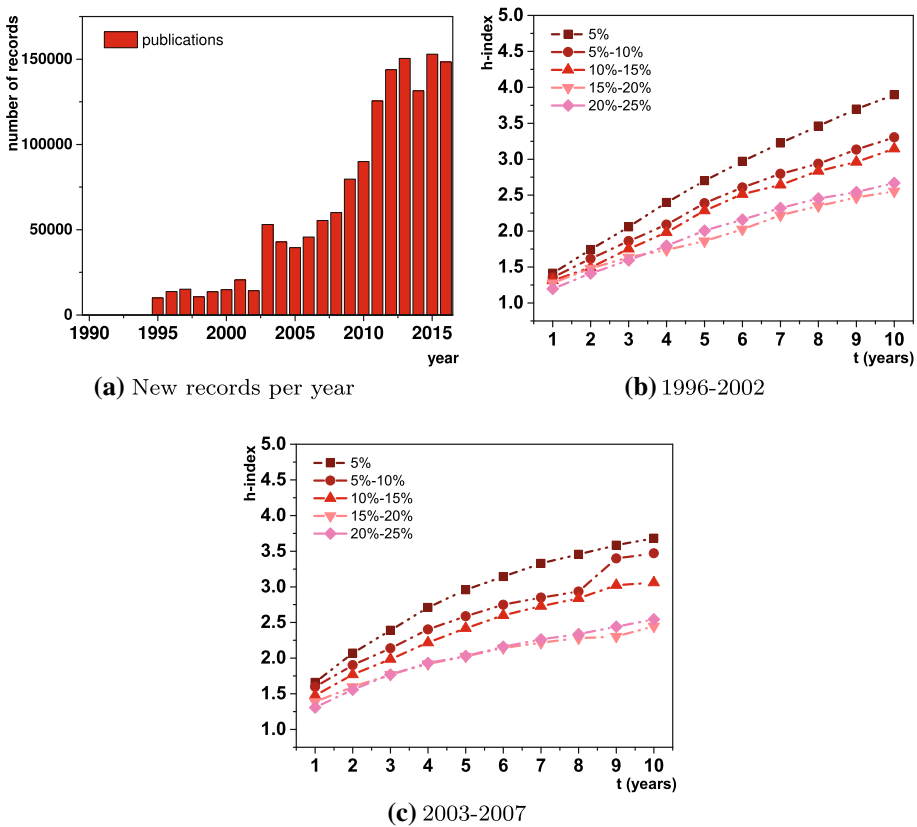
are always below 0.1. For others, their values are all above 0.1, even the maximum value can reach 0.28. It indicates that most *Top<sub>10</sub>* advisees are coached by *Top<sub>10</sub>* advisors.

Selective bias is a possible explanation for these phenomena. The advisor–advisee relationship is a two-way choice relationship, quality advisees tend to seek top advisors and vice versa. Moreover, we have also found that the gap between these advisees is closely related to the advisors’ academic age and the time duration of their collaboration. In the first three years of collaboration, the gap between advisees is relatively small. Since the fourth year, the gap begins to grow wider.

### Publication rates analysis

As mentioned before, we have separated advisors according to their AA in the experiment. However, considering the publication rates increased over time, it may have a biased effect





**Fig. 9** Advisees' h-indices whose advisors' AA is 20 in different time periods, as well as their h-indices are ranking at the top 5%, 5%–10%, 10%–15%, 15%–20%, 20%–25%. **a** The number of new entries in the database per year. **b** Advisees' average h-index mentored by advisors whose AA is 20 in 1996–2002. **c** Advisees' average h-index mentored by advisors whose AA is 20 in 2003–2007

on the scholars' number of publications, citations, and h-indices. So we carry out the experiment to verify whether the publications rates impact our findings.

Figure 9a displays the number of new entries in the database per year. From the data in Fig. 9a, it is apparent that the growth rate of the publication is different at different stages. The number of new records increases sharply in 2003 and 2011. To be more specific, there are less than 25,000 new publications per year in 1999–2002 but over 50,000 in 2003–2007, almost two times more than the first period. In order to ensure all of the advisees have plenty of time to accumulate enough publications, we choose advisors whose academic age is 20 in 1996–2002 and 2003–2007, respectively. We compare the average academic performance (h-index) of their advisees. In Fig. 9b, and c, the polylines show the advisees' average h-index mentored by different h-index ranking advisors. Data in these figures can be compared with the data in Fig. 5, which is the basis of our findings. A positive correlation can be found between advisors' and advisees' academic performance. So the publication rates have limited influence on the phenomena we have found in this work.

## Discussion

Moreover, we have processed our dataset and do experiments on the part of the real dataset to explore the correlation between the advisor and their academic grandchildren. We extract 4256 advisor–advisee pairs (645 advisors) and compile the statistics on the data. For advisors, we still divide them into two groups according to their h-indices: Top<sub>10</sub> advisors and *Res* advisors. And then we analyze their grandchildren's impact. We discover that the average h-index of Top<sub>10</sub> advisors' grand-advisees is 40% higher than the grand-advisees mentored by *Res* advisors. It illustrates that there may be also a correlation existing between them and it needs in-depth research. Usually, PhD students are guided directly by their advisors, thus we only focus on the direct advisor–advisee relationships. Through this analysis, a comparatively deep analysis should also be made on the relationship between the advisor and their academic grandchild. We will carry out an in-depth study of this issue in future work.

There are still a few limitations in this work. Firstly, the conclusions are efficacious only for Computer Science, which is a rapidly developing discipline and has the characteristic of spreading its knowledge in conferences. It would be interesting to explore the anatomy of the relationship in other traditional areas of knowledge if we can get the large-scale advisor–advisee relationships dataset for the discipline. Secondly, we only show the correlation between advisors' academic characteristics and advisees' academic performance. Other factors may also produce these phenomena, such as selective bias and the rank of institutions. It is necessary to further explore the cause of these phenomena. Thirdly, it is unfortunate that the study only illustrates the advisors' benefits derived from advisees on the formal relations. In terms of relevant experience, the informal advisor–advisee relationship is also an important part of mentorship. In future work, through the continuous improvement of the dataset, we will conduct the in-depth study on the informal advisor–advisee relationships.

## Conclusion

This study presents advisees' potential academic performance of choosing different advisors with different academic ages or academic levels in the field of Computer Science. It has identified the relationship between advisors' academic age and advisees' performance, which experiences an initial growth, follows a sustaining stage, and finally ends up with a declining trend. The second major finding is that accomplished advisors can bring up skilled advisees, which is evidenced by advisees' academic performance advised by different advisors. Taken together, these findings suggest a significant role for advisors in promoting their advisees. This research extends our knowledge of scholars' career success and will serve as a base for future studies.

In summary, through the analysis of the relationship between advisors and advisees, the findings have significant implications to understand the relationship between advisors' academic characteristics and advisees' performance. Moreover, it can shed light on the advisor recommendation. However, it should be noted that the results may differ in different research fields, it would be interesting to explore the anatomy of the relationships in other areas of knowledge.

**Acknowledgements** The authors extend their appreciation to the Deanship of Scientific Research at King Saud University for funding this work through research group NO (RG-1438-027). Xiangjie Kong is

supported by Fundamental Research Funds for the Central Universities under Grant NO (DUT18JC09), and China Scholarship Council under Grant NO (201706060067).

## References

- Amjad, T., Ding, Y., Xu, J., Zhang, C., Daud, A., Tang, J., et al. (2017). Standing on the shoulders of giants. *Journal of Informetrics*, 11(1), 307–323.
- Anderson, C. (2006). *The long tail: Why the future of business is selling less of more*. New York: Hachette Books.
- Azoulay, P. (2012). Research efficiency: Turn the scientific method on ourselves. *Nature*, 484(7392), 31–32.
- Borders, L. D., Wester, K. L., Granello, D. H., Chang, C. Y., Hays, D. G., Pepperell, J., et al. (2012). Association for counselor education and supervision guidelines for research mentorship: Development and implementation. *Counselor Education and Supervision*, 51(3), 162–175.
- Bozionelos, N., Bozionelos, G., Polychroniou, P., & Kostopoulos, K. (2014). Mentoring receipt and personality: Evidence for non-linear relationships. *Journal of Business Research*, 67(2), 171–181.
- Chao, G. T. (1997). Mentoring phases and outcomes. *Journal of Vocational Behavior*, 51(1), 15–28.
- Chao, G. T., Walz, P., & Gardner, P. D. (1992). Formal and informal mentorships: A comparison on mentoring functions and contrast with nonmentored counterparts. *Personnel Psychology*, 45(3), 619–636.
- Dobson, L. (2013). Effective practices of formal mentoring programs. *Math Alliance Research Study*, 3, 1–3.
- Fagenson, E. A. (1989). The mentor advantage: Perceived career/job experiences of protégés versus non-protégés. *Journal of Organizational Behavior*, 10(4), 309–320.
- Florea, L., Cheung, Y. H., & Herndon, N. C. (2013). For all good reasons: Role of values in organizational sustainability. *Journal of Business Ethics*, 114(3), 393–408.
- Ghosh, R., & Reio, T. G. (2013). Career benefits associated with mentoring for mentors: A meta-analysis. *Journal of Vocational Behavior*, 83(1), 106–116.
- Glänzel, W., Heffer, S., & Thijs, B. (2017). Lexical analysis of scientific publications for nano-level scientometrics. *Scientometrics*, 111, 1–10.
- Gollapalli, S. D., Mitra, P., & Giles, C. L. (2011). Ranking authors in digital libraries. In *Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries*, (pp. 251–254). ACM.
- Hu, C., Wang, S., Yang, C.-C., & Wu, T.-Y. (2014). When mentors feel supported: Relationships with mentoring functions and protégés' perceived organizational support. *Journal of Organizational Behavior*, 35(1), 22–37.
- Johnson, W. B., & Ridley, C. R. (2015). *The elements of mentoring*. Basingstoke: Macmillan.
- Kram, K. E. (1983). Phases of the mentor relationship. *Academy of Management Journal*, 26(4), 608–625.
- Langfeldt, L., Benner, M., Sivertsen, G., Kristiansen, E. H., Aksnes, D. W., Borlaug, S. B., et al. (2015). Excellence and growth dynamics: A comparative study of the Matthew effect. *Science and Public Policy*, 42, scu083.
- Lehmann, S., Jackson, A. D., & Lautrup, B. E. (2006). Measures for measures. *Nature*, 444(7122), 1003–1004.
- Letchford, A., Moat, H. S., & Preis, T. (2015). The advantage of short paper titles. *Royal Society open science*, 2(8), 150266.
- Ley, M. (2009). DBLP: Some lessons learned. *Proceedings of the VLDB Endowment*, 2(2), 1493–1500.
- Malmgren, R. D., Ottino, J. M., & Amaral, L. A. N. (2010). The role of mentorship in protégé performance. *Nature*, 465(7298), 622–626.
- Moreira, C., Calado, P., & Martins, B. (2011). Learning to rank for expert search in digital libraries of academic publications. In *Portuguese conference on artificial intelligence*, (pp. 431–445). Springer.
- Murphy, T. H. (2015). On great teachers. *Journal of Legal Studies Education*, 32(1), 223–227.
- Newman, M. E. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98(2), 404–409.
- Owens, B. (2013). Research assessments: Judgement day. *Nature*, 502(7471), 288–290.
- Scandura, T. A., & Ragins, B. R. (1993). The effects of sex and gender role orientation on mentorship in male-dominated occupations. *Journal of Vocational Behavior*, 43(3), 251–265.
- Sinatra, R., Wang, D., Deville, P., Song, C., & Barabási, A.-L. (2016). Quantifying the evolution of individual scientific impact. *Science*, 354(6312), aaf5239.
- Singh, R., Ragins, B. R., & Tharenou, P. (2009). Who gets a mentor? A longitudinal assessment of the rising star hypothesis. *Journal of Vocational Behavior*, 74(1), 11–17.

- Tang, J., Jin, R., & Zhang, J. (2008). A topic modeling approach and its integration into the random walk framework for academic search. In *Eighth IEEE international conference on data mining, 2008. ICDM'08*, (pp. 1055–1060). IEEE.
- Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., & Su, Z. (2008). Arnetminer: Extraction and mining of academic social networks. In *Proceedings of the 14th ACM SIGKDD international conference on knowledge discovery and data mining*, (pp. 990–998). ACM.
- Tuesta, E. F., Delgado, K. V., Mugnaini, R., Digiampietri, L. A., Mena-Chalco, J. P., & Pérez-Alcázar, J. J. (2015). Analysis of an advisor–advisee relationship: An exploratory study of the area of exact and earth sciences in Brazil. *PloS ONE*, *10*(5), e0129065.
- Wanberg, C. R., Kammeyer-Mueller, J., & Marchese, M. (2006). Mentor and protégé predictors and outcomes of mentoring in a formal mentoring program. *Journal of Vocational Behavior*, *69*(3), 410–423.
- Wang, W., Liu, J., Xia, F., King, I., & Tong, H. (2017). Shifu: Deep learning based advisor–advisee relationship mining in scholarly big data. In *Proceedings of the 26th international conference on world wide web companion*, (pp. 303–310). International World Wide Web Conferences Steering Committee.
- Wang, W., Liu, J., Yu, S., Zhang, C., Xu, Z., & Xia, F. (2016). Mining advisor–advisee relationships in scholarly big data: A deep learning approach. In *Proceedings of the 16th ACM/IEEE-CS on joint conference on digital libraries*.
- Wang, W., Yu, S., Bekele, T. M., Kong, X., & Xia, F. (2017). Scientific collaboration patterns vary with scholars academic ages. *Scientometrics*, *112*(1), 329–343.
- Xia, F., Wang, W., Bekele, T. M., & Liu, H. (2017). Big scholarly data: A survey. *IEEE Transactions on Big Data*, *3*(1), 18–35.
- Young, A. M., & Perrewe, P. L. (2000). What did you expect? An examination of career-related support and social support among mentors and protégés. *Journal of Management*, *26*(4), 611–632.